Molecular Quantum Similarity Measures as an Alternative to Log P Values in QSAR Studies

LLUÍS AMAT,[1] RAMON CARBÓ-DORCA,[1] ROBERT PONEC[2]

[1] *Institute of Computational Chemistry, University of Girona, Catalonia 17071, Spain*
[2] *Institute of Chemical Process Fundamentals, Czech Academy of Sciences, Prague, Czech Republic*

Received 13 March 1998; accepted 31 May 1998

ABSTRACT: A new molecular descriptor of hydrophobicity based on molecular quantum similarity measures (MQSM), which can be used to replace the log P parameter in QSAR studies, is proposed. Unlike the majority of existing approaches for calculation of log P, the present methodology does not rely on the use of fragment additive contributions, but rather it is based on the comparison of quantum chemically calculated electron density distributions of a given molecule in water and in 1-octanol, using MQSM. The method has been tested on a broad series of 58 molecules including such structural types as aliphatic hydrocarbons, alcohols, amines, halides, carboxylic acids, esters, amides, and ketones, as well as more complex systems with two functional groups. In all cases investigated, an excellent linear relationship between calculated MQSM and log P values was found. Additionally, an example of QSAR analysis is presented using MQSM instead of log P values, corresponding to predict the narcosis of tadpoles. © 1998 John Wiley & Sons, Inc. J Comput Chem 19: 1575–1583, 1998

Keywords: 1-octanol/water partition coefficient; molecular quantum similarity measures; atomic shell approximation; promolecular densities; QSAR

Correspondence to: R. Carbó-Dorca e-mail: director@iqc.udg.es

Contract/grant sponsors: NATO; Grant Agency of the Czech Republic, contract/grant number: 203/95/0650; contract/grant sponsor: CICYT, contract/grant number: SAF 96-0158.

CCC 0192-8651 / 98 / 141575-09

Introduction

Because of the complex interactions between biologically active drugs and the corresponding receptors, an important step in the design of new active compounds still belongs to the so-called QSAR approach, where various empirical relationships are used to correlate experimental biological data with various molecular descriptors. Among the most commonly used descriptors, it has been proven that one is especially universal in these correlation tasks: the molecular partition coefficient in 1-octanol/water system. The so-called ''log P'' is a simple measure of the hydrophobic/lipophilic character of the drug.

Although this parameter is generally available from experimental determinations for extensive series of molecules, there are situations in which its values are still unknown. This may occur, for example, when the studied molecules are completely new or when the experimental determination is difficult. For that reason, and because of the importance of this parameter, several evaluation methods have been proposed.

The log P computation techniques are usually based on additive group contributions. The first of such procedures was proposed by Fujita et al.[1] In this approach, log P is estimated for a given chemical structure as a sum of the log P of the parent solute plus a π term, which is defined as the difference of log P between a particular substituent and the replaced hydrogen atom. Another way of calculating log P is the fragment method, introduced by Rekker and coworkers.[2,3] In this case, log P values are constructed from hydrophobic fragment constants. Later, Hansch and Leo[4] proposed new tables for fragment values, where some correction terms were included, and some novel definitions, like fundamental fragments, isolating and nonisolating carbons, were introduced. In addition to these methods, another procedure was reported by Klopman and coworkers in this journal.[5,6] They based their approach on the idea of hydrophobic atomic contributions. In this way, they developed a procedure in which the effect of actual charge densities on the hydrophobicity was explicitly considered. More recently, Ghose and Crippen,[7] using a least-squares technique and a set of 494 molecules with known log P, classified the atoms H, C, O, N, S, and halogens into 90 different types, and calculated their hydrophobic atomic contributions to log P.

In the present study, an alternative to log P parameter values in QSAR studies, based on molecular quantum similarity measures (MQSM), is presented. Using this approach, the interplay between the hydrophobic and lipophilic nature of the molecule, characterized experimentally by log P values, is deduced from the comparison of calculated electron density distributions of the molecule in water and 1-octanol. The approach has been applied to a broad series of 58 molecules including various structural types; for instance, hydrocarbons, alcohols, amines, carboxylic acids, halides, esters, ketones, amides, amino acids, and chlorhydrines. In all cases, an almost perfect linear relationship between calculated similarity measures and log P values has been found. In addition, a QSAR example is presented for prediction of the narcosis of tadpoles.

Theoretical

The philosophy underlying the introduction and the application of MQSM was thoroughly reported in previous studies.[8–17] Here it is considered worthwhile to recapitulate briefly the basic theoretical ideas for better understanding the present discussion. The application of MQSM to quantitative structure–property relationships (QSPR)[18–20] is based on the idea that properties of molecules are determined, even if in a very complex way, by their electronic structure. Therefore, the regularities in molecular properties can be studied using similarity measures between molecular electronic structures. The simplest quantum-chemical quantity, characterized from the point of view of electronic distribution of molecular structure, is the first-order density function, $\rho(\mathbf{r})$, and, as it will be shown in what follows, MQSM are generally derived from the pairwise comparison of these densities for a series of studied molecules. In this way, the MQSM of two molecules A and B can generally be defined as

$$Z_{AB}(\Omega) = \int\int \rho_A(\mathbf{r}_1)\Omega(\mathbf{r}_1,\mathbf{r}_2)\rho_B(\mathbf{r}_2)\, d\mathbf{r}_1\, d\mathbf{r}_2 \quad (1)$$

Depending on the actual choice of two-electron operator, Ω, various MQSMs can be obtained. In this study, the simplest possibility is used, with the weighting operator, Ω, given by the Dirac

delta function, $\Omega(\mathbf{r}_1, \mathbf{r}_2) = \delta(\mathbf{r}_1 - \mathbf{r}_2)$, which results in the so-called overlap-like MQSM simplification of the [eq. (1)] integral

$$Z_{AB} = \int \rho_A(\mathbf{r})\,\rho_B(\mathbf{r})\,d\mathbf{r} \tag{2}$$

It is important to point out that the values of the integral [eq. (2)] depend on the actual position and mutual orientation of molecules A and B, so that the optimization of this position yielding maximum similarity is generally required.[21] In addition, the large number of four-center integrals, which must be computed within the usual LCAO approximation, makes the calculation of the integral very cumbersome. To reduce computation time, a simplified approach was proposed some time ago, known as the *atomic shell approximation* (ASA).[22, 23] Using this approach, the molecular electron density is fitted in a least-squares manner to the linear combination of spherical functions centered on individual atoms.

Moreover, to simplify the calculations even further, a promolecular approximation[23] is used, in which the molecular density function is constructed as simple sums of atomic density functions

$$\rho_A^{ASA}(\mathbf{r}) = \sum_{a \in A} P_a\,\rho_a^{ASA}(\mathbf{r}) \tag{3}$$

In this work, the coefficient, P_a, has been defined as the valence electron density on atom a. The atomic density functions are constructed, using ASA formalism, as a linear combination of normalized 1S-type GTO

$$\rho_a^{ASA}(\mathbf{r}) = \sum_{i \in a} w_i |S_i(\mathbf{r} - \mathbf{R}_a; \zeta_i)|^2 \tag{4}$$

where the coefficients, w_i, fulfill the convex constraints

$$\{w_i > 0, \forall i\} \wedge \left\{ \sum_i w_i = 1 \right\}$$

to preserve the statistical distribution probability meaning of the approximate density function.[24] Both sets of coefficients $\{w_i\}$ and exponents $\{\zeta_i\}$ have been optimized to minimize the quadratic error between HF/3-21G atomic density function and the ASA density function.[23] These sets of atomic parameters can be downloaded from a www site.[25] In this work, the rule used to generate promolecular density functions is one function for H; three functions for C, N and O; and four functions for Cl.

Computational Details

Having introduced the basic MQSM formalism, one can proceed to the practical exploitation of these measures for the calculation of parameters that can replace the log P values in QSAR studies. As previously noted, the partition coefficient between water and 1-octanol is a quantity often used in QSAR as a descriptor of hydrophobicity of the molecule. Thus, in this descriptor, the complex phenomenon affected by many not completely known factors is reduced to a simple single number. Although the factors influencing log P values are certainly difficult to specify, the relative solubility of the molecule in polar water and nonpolar 1-octanol is, certainly, one of the most important. Based on this assumption, Klopman and coworkers proposed to connect this solubility to the electrostatic interactions involving charge densities and postulated that the partition coefficient is primarily influenced only by these interactions.[5, 6] The basic idea of the present approach arises from this proposal by Klopman, but instead of using discrete point charges on individual atoms, here a continuous density function, $\rho(\mathbf{r})$, is chosen to characterize the charge distribution in the molecule. Taking into account that the electron density function, ρ_A^X, of molecule A, immersed in solvent X, is generally affected by the solvent and differs from the gas phase density, $\rho_A(\mathbf{r})$, then the relative solubility of the molecule in two different solvents, X and Y, could presumably be related to the similarity of the molecular electron density functions, $\rho_A^X(\mathbf{r})$ and $\rho_A^Y(\mathbf{r})$, in these solvents. This similarity measure can quantitatively be characterized by the value of a specific form of the integral [eq. (2)]

$$Z_{AA}^{XY} = \int \rho_A^X(\mathbf{r})\,\rho_A^Y(\mathbf{r})\,d\mathbf{r} \tag{5}$$

Using *promolecular* ASA density functions, this overlap-like MQSM is simplified to

$$Z_{AA}^{XY} = \sum_{a \in A} \sum_{b \in A} P_a^X P_b^Y Z_{ab} \tag{6}$$

where P_a and P_b correspond to the same corrected atomic charges as defined in eq. (3), and the atomic quantum similarity contributions Z_{ab} are calcu-

lated as the integrals

$$Z_{ab} = \sum_{i \in a} \sum_{j \in b} w_i w_j \int |S_i(\mathbf{r} - \mathbf{R}_a)|^2 |S_j(\mathbf{r} - \mathbf{R}_b)|^2 \, d\mathbf{r} \tag{7}$$

which correspond to a simple and well-known overlap between two 1S-GTO functions.[26]

Before calculating the theoretical descriptor, Z_{AA}^{XY}, a complete optimization of geometry at the ab initio HF level of theory using the 3-21G* basis set and the Gaussian-94 program[27] has been performed. The most stable molecular conformation has been used for the calculation of Z_{AA}^{XY}. Such calculations have consisted of the explicit inclusion of solvent effect on molecular electron density within the so-called polarizable continuum (PCM) model[28,29] included in the Gaussian program. Within this approach, a molecule with fixed, gas-phase-optimized geometry is placed in a cavity surrounded by a medium characterized by a dielectric constant, and the effect of this polarizable medium on electron distribution in a molecule is explicitly evaluated. The values for dielectric constants of $\varepsilon = 80.4$ and $\varepsilon = 10.3$[30] have been used for water and 1-octanol solvents, respectively. The same model has already been used to study, by means of MQSM, the effect of solvation on molecular charge distributions.[31] The ASA electron density functions of a given molecule A in these two solvents, $\rho_A^{w}(\mathbf{r})$ and $\rho_A^{o}(\mathbf{r})$, are constructed using the gas-phase-optimized geometry and the valence atomic charges determined from the solvent calculations. Finally, the MQSM Z_{AA}^{wo} is computed using these density functions and the integral defined in eq. (5). Here it should be stressed that, although the values of similarity measures, Z_{AB}, generally depend on the mutual position of the molecules so that the optimization of this position is generally required, in the case of self-similarity measures, Z_{AA}, the situation is much simpler, because, for identical molecules, the problem of optimal superposition of corresponding structures is trivial.

Results

LINEAR RELATIONSHIPS BETWEEN LOG P VALUES AND SIMILARITY MEASURES

Having presented the basic methodological background, the results of the practical application of the aforementioned approach will be given. To test the new methodology, an extensive series of 58 molecules involving various structural types has been tested. In all cases the most stable molecular geometries have been considered. These structural types involve a series of simple monofunctional derivatives like hydrocarbons R—H, alcohols R—OH, halides R—Cl, amines R—NH_2, carboxylic acids R—COOH, esters R—$COOCH_3$, amides H—CONHR, R—$CONH_2$, and ketones R—CO—R′ as well as representatives of more complex molecules with polyfunctional groups [chlorhydrines Cl—$(CH_2)_n$—OH, and amino acids R—$CH(NH_2)$—COOH]. The values of log P and the similarity measures, Z_{AA}^{wo}, characterizing, in each particular case, the resemblance of the electron density functions in water and 1-octanol, are described in Table I. The results of the linear relationships between calculated MQSMs and log P values are summarized in Table II and for several representative series also in graphical form in Figures 1–3.

APPLICATION OF MQSM IN QSAR STUDIES

The dependence of log P values on the calculated theoretical MQSM descriptor, Z_{AA}^{wo}, is indeed excellent, and similarly good correlations are observed in all molecular series investigated. This result is very important because numerous biological data exist for which the log P is the only decisive parameter in the corresponding QSAR model. An example summarizing such empirical QSAR studies may be found in ref. 32, in which an extensive comparison of diverse biological effects of simple molecules was performed. In view of the excellent correlation just indicated, it is possible to expect that the proposed theoretical descriptor could be used to replace the log P parameter in many empirical QSAR cases. As an example, Table III shows a comparison of experimental and calculated biological activities of aliphatic alcohols, ketones, and acetic acid esters for the narcosis of tadpoles. However, in connection with the linear relationships in Table II, successful correlations with the theoretical descriptor Z_{AA}^{wo} exist only within a series of structurally related molecules. The success of the computed linear relationships is due to the close structural similarity of the series of molecules investigated. In this respect, the simplest situation is with the monofunctional aliphatic derivatives CH_3—$(CH_2)_n$—X, in which the only variable factor is the number of CH_2 groups in the alkyl chain. As a consequence of this simplicity it would be possible to also expect correlations of log P values with numbers of CH_2 groups, or,

TABLE I.
Log *P* Values and Calculated MQSM Z_{AA}^{WO} for a Series of Selected Molecules.

Molecule	Formula	Log P	Z_{AA}^{WO}
Hydrocarbons (R—H)			
1 Methane	CH_4	1.09[a]	20.38
2 Ethane	CH_3CH_3	1.81[a]	37.47
3 Propane	$CH_3CH_2CH_3$	2.36[a]	54.53
4 Butane	$CH_3CH_2CH_2CH_3$	2.89[a]	71.63
5 Pentane	$CH_3CH_2CH_2CH_2CH_3$	3.39[a]	88.73
6 Hexane	$CH_3CH_2CH_2CH_2CH_2CH_3$	3.90[a]	105.66
Amines ($R—NH_2$)			
7 Methylamine	CH_3NH_2	−0.57[a]	53.54
8 Ethylamine	$CH_3CH_2NH_2$	−0.13[a]	70.82
9 Propylamine	$CH_3CH_2CH_2NH_2$	0.48[a]	88.06
10 Butylamine	$CH_3CH_2CH_2CH_2NH_2$	0.97[a]	105.03
Primary alcohols (R—OH)			
11 Methanol	CH_3OH	−0.77[a]	73.69
12 Ethanol	CH_3CH_2OH	−0.31[a]	91.23
13 Propanol	$CH_3CH_2CH_2OH$	0.25[a]	108.49
14 Butanol	$CH_3CH_2CH_2CH_2OH$	0.88[a]	125.50
15 Pentanol	$CH_3CH_2CH_2CH_2CH_2OH$	1.56[a]	142.30
16 Hexanol	$CH_3CH_2CH_2CH_2CH_2CH_2OH$	2.03[a]	159.68
Secondary alcohols ($R_1—CH(OH)—R_2$)			
17 2-Propanol	$(CH_3)_2CHOH$	0.05[a]	108.30
18 2-Butanol	$CH_3CH_2CH(OH)CH_3$	0.61[a]	125.60
19 2-Pentanol	$CH_3CH_2CH_2CH(OH)CH_3$	1.19[a]	142.79
20 3-Pentanol	$CH_3CH_2CH(OH)CH_2CH_3$	1.21[a]	142.51
21 2-Hexanol	$CH_3CH_2CH_2CH_2CH(OH)CH_3$	1.76[a]	159.48
22 3-Hexanol	$CH_3CH_2CH_2CH(OH)CH_2CH_3$	1.65[a]	159.49
Ketones (R_1COR_2)			
23 Acetone	CH_3COCH_3	−0.24[a]	105.44
24 2-Butanone	$CH_3COCH_2CH_3$	0.29[a]	122.41
25 2-Pentanone	$CH_3COCH_2CH_2CH_3$	0.91[a]	139.58
26 3-Pentanone	$CH_3CH_2COCH_2CH_3$	0.99[a]	139.85
27 2-Hexanone	$CH_3COCH_2CH_2CH_2CH_3$	1.38[a]	156.78
28 3-Hexanone	$CH_3CH_2COCH_2CH_2CH_3$	1.45[b]	156.85
Esters (CH_3COOR)			
29 Acetic acid	CH_3COOH	−0.17[a]	141.36
30 Methylacetate	CH_3COOCH_3	0.18[a]	157.71
31 Ethylacetate	$CH_3COOCH_2CH_3$	0.73[a]	175.18
32 Propylacetate	$CH_3COOCH_2CH_2CH_3$	1.24[a]	192.41
33 Butylacctate	$CH_3COOCH_2CH_2CH_2CH_3$	1.78[a]	209.31
Carboxylic acids (RCOOH)			
29 Acetic acid	CH_3COOH	−0.17[a]	141.36
34 Propionic acid	CH_3CH_2COOH	0.33[a]	159.10
35 Butyric acid	$CH_3CH_2CH_2COOH$	0.79[a]	175.93
36 Valeric acid	$CH_3CH_2CH_2CH_2COOH$	1.39[a]	193.16
37 Hexanoic acid	$CH_3CH_2CH_2CH_2CH_2COOH$	1.92[a]	209.78
Amides (HCONHR)			
38 Formamide	$HCONH_2$	−1.51[a]	104.55
39 *N*-methylformamide	$HCONHCH_3$	−0.97[a]	120.65
40 *N*-ethylformamide	$HCONHCH_2CH_3$	−0.43[b]	138.17
41 *N*-propylformamide	$HCONHCH_2CH_2CH_3$	0.11[b]	155.55
42 *N*-butylformamide	$HCONHCH_2CH_2CH_2CH_3$	0.65[b]	172.39

TABLE I.
(Continued)

Molecule	Formula	Log P	Z_{AA}^{wo}
Amides ($RCONH_2$)			
43 Acetamide	CH_3CONH_2	−1.26[a]	123.03
44 Propionamide	$CH_3CH_2CONH_2$	−0.66[a]	140.23
45 Butyramide	$CH_3CH_2CH_2CONH_2$	−0.21[a]	157.41
46 Valeramide	$CH_3CH_2CH_2CH_2CONH_2$	0.33[b]	174.44
47 Caproamide	$CH_3CH_2CH_2CH_2CH_2CONH_2$	0.87[b]	191.32
Chlorides (R—Cl)			
48 Methylchloride	CH_3Cl	0.91[a]	191.48
49 Ethylchloride	CH_3CH_2Cl	1.43[a]	209.33
50 1-Chloropropane	$CH_3CH_2CH_2Cl$	2.04[a]	226.58
51 1-Chlorobutane	$CH_3CH_2CH_2CH_2Cl$	2.64[a]	243.82
Alcohols, chlorides ($Cl—(CH_2)_n—OH$)			
52 2-Chloroethanol	CH_2OHCH_2Cl	−0.06[a]	262.09
53 3-Chloro-1-propanol	$CH_2OHCH_2CH_2Cl$	0.50[a]	279.57
54 4-Chloro-1-butanol	$CH_2OHCH_2CH_2CH_2Cl$	0.85[a]	297.44
Amino acids ($R—CH(NH_2)COOH$)			
55 Glycine	CH_2NH_2COOH	−3.21[a]	175.05
56 Alanine	CH_3CHNH_2COOH	−2.96[a]	192.30
57 α-Aminobutyric	$CH_3CH_2CHNH_2COOH$	−2.53[a]	209.08
58 α-Aminovaleric	$CH_3CH_2CH_2CHNH_2COOH$	−2.11[a]	225.88

[a] From ref. 33.
[b] From ref. 32.

more or less equivalently, with the total number of electrons (N) in the molecule.

BEHAVIOR OF CORRELATION LINES

In this subsection, conclusions are deduced from the results presented. A point worthy of mention concerns the fact that the slopes of the linear relationships between log P and Z_{AA}^{wo} for all series of monofunctional derivatives are almost the same. This result is not at all trivial, and it is in fact this near equivalence of slopes that constitutes the basis of the simple Fujita et al. method of log P calculation.[1] Another interesting aspect regarding the close resemblance of the slopes of log P vs. Z_{AA}^{wo} correlations is based on the fact that, while

TABLE II.
Calculated Statistical Parameters of Linear Relationships between MQSM Z_{AA}^{wo} and Log P Values for the Series of Molecules Described in Table I.

Series	Linear equation	*n*	r^2
R—H	$\log P = 0.0323 \times Z_{AA}^{wo} + 0.5332$	6	0.996
$R—NH_2$	$\log P = 0.0305 \times Z_{AA}^{wo} - 2.2295$	4	0.996
R—OH	$\log P = 0.0337 \times Z_{AA}^{wo} - 3.3314$	6	0.996
$R_1—CH(OH)—R_2$	$\log P = 0.0324 \times Z_{AA}^{wo} - 3.4537$	6	0.996
R—OH and $R_1—CH(OH)—R_2$	$\log P = 0.0311 \times Z_{AA}^{wo} - 3.1409$	12	0.972
R_1COR_2	$\log P = 0.0326 \times Z_{AA}^{wo} - 3.6617$	6	0.993
CH_3COOR	$\log P = 0.0291 \times Z_{AA}^{wo} - 4.3443$	5	0.996
RCOOH	$\log P = 0.0307 \times Z_{AA}^{wo} - 4.5390$	5	0.998
HCONHR	$\log P = 0.0317 \times Z_{AA}^{wo} - 4.8064$	5	0.999
CH_3COOR, RCOOH, and HCONHR	$\log P = 0.0316 \times Z_{AA}^{wo} - 4.7706$	14	0.996
$RCONH_2$	$\log P = 0.0307 \times Z_{AA}^{wo} - 5.0208$	5	0.999
R—Cl	$\log P = 0.0333 \times Z_{AA}^{wo} - 5.4906$	4	0.998
$Cl—(CH_2)_n—OH$	$\log P = 0.0257 \times Z_{AA}^{wo} - 6.7644$	3	0.981
$R—CH(NH_2)COOH$	$\log P = 0.0220 \times Z_{AA}^{wo} - 7.1191$	4	0.986

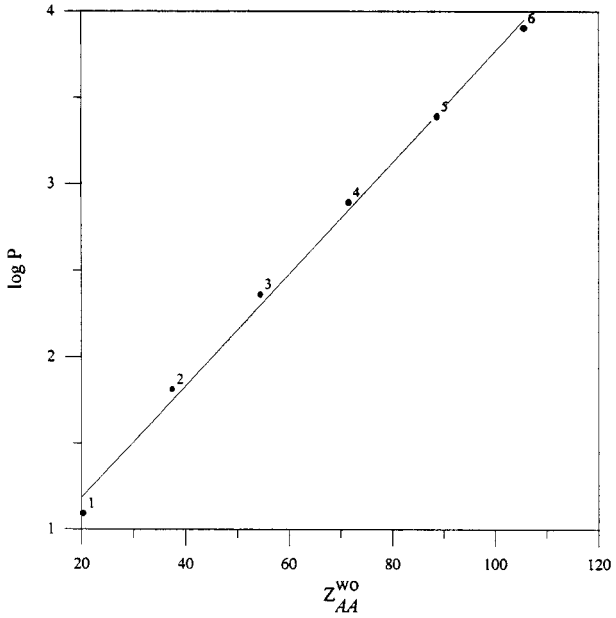

FIGURE 1. Dependence of calculated MQSM for a series of hydrocarbons (R—H) on log P values.

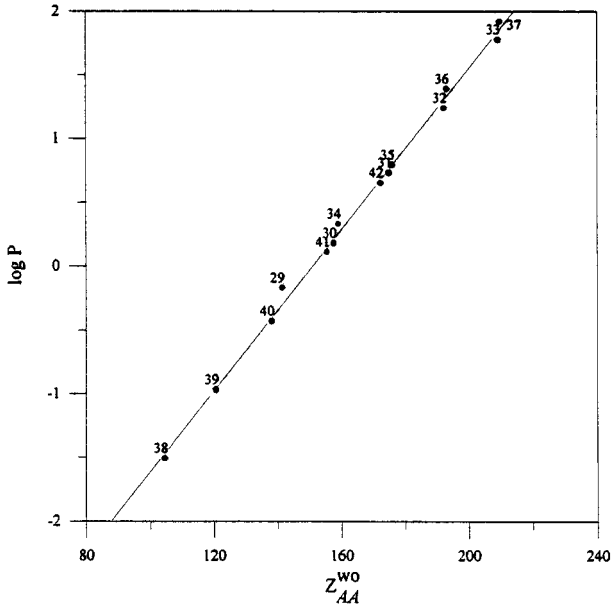

FIGURE 3. Dependence of calculated MQSM for a series of acetic acid esters (CH_3COOR), carboxylic acids (RCOOH), and amides (HCONHR) on log P values.

individual correlation lines are generally observed for each particular class of molecules, the chemical similarity of functional groups can also been seen in the similarity of intercepts. As a consequence, it is possible to combine the data for these groups to determine a single, common linear relationship; for example, in cases dealing with primary and secondary alcohols as well as for esters, carboxylic acids, and amides H—CONHR (see Table II). However, a more complex situation appears with molecules containing two or more functional groups, where the existence of linear relationships

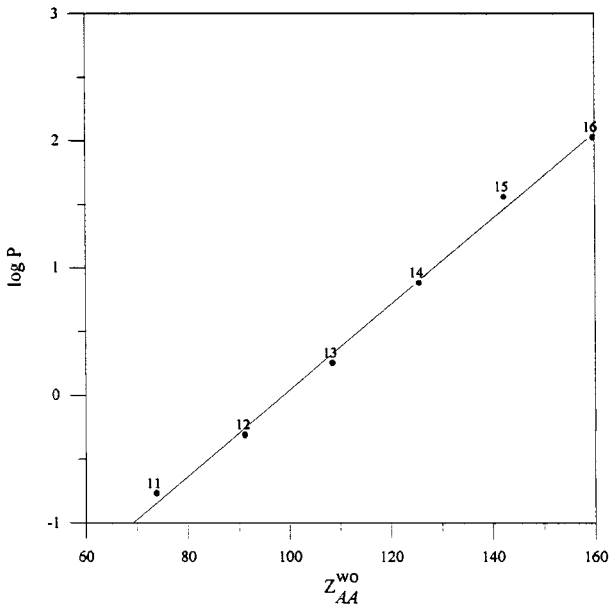

FIGURE 2. Dependence of calculated MQSM for a series of primary alcohols (R—OH) on log P values.

TABLE III.
Comparison of Experimental and Calculated Biological Activities of Aliphatic Alcohols, Ketones, and Acetic Acid Esters in Narcosis of Tadpoles.

Molecule	Log 1 / C	
	Obs.[a]	Calc.[b]
Methanol	0.24	0.19
Ethanol	0.54	0.58
1-Propanol	0.96	0.97
1-Butanol	1.42	1.35
2-Propanol	0.89	0.96
Acetone	0.54	0.54
2-Butanone	1.04	1.04
3-Pentanone	1.54	1.54
Methylacetate	1.10	1.11
Ethylacetate	1.52	1.52
Propylacetate	1.96	1.93
Butylacetate	2.30	2.32

[a] From ref. 34.
[b] Calculated using the equations: log 1 / C = 0.0225 Z_{AA}^{WO} − 1.4713 for aliphatic alcohols; log 1 / C = 0.0291 Z_{AA}^{WO} − 2.5217 for ketones; and log 1 / C = 0.0235 Z_{AA}^{WO} − 2.5942 for acetic acid esters.

between log P and Z_{AA}^{wo} is also reported, but the slopes of corresponding correlations are considerably different (see Table II). This result is of interest because it demonstrates that deviations from simple group additivity schemes of log P calculation can be expected in these cases. This is not a surprising finding, because, in cases of complex structural series, various correcting terms are required to improve the accuracy of simple additivity schemes. Thus, the just-reported sensitivity of the slopes of log P vs. correlations in cases of multiple substituted reaction series brings a clear theoretical rationale for inclusion of such corrections.

In summarizing the results presented it is perhaps possible to conclude that, in spite of its simplicity, the present method provides an improved approach for computation of log P values and allows their substitution by using MQSM in a completely general way. Although the method is preferably applicable to closely structurally related series of molecules, the accuracy of the predictions is, in these cases, generally higher than for other existing methods. This can best be shown by comparing our calculated log P values with predictions obtained by other alternative methods (see Tables IV and V).

Conclusions

During the past few years a systematic effort was devoted in our laboratory to the theoretical rationalization of QSPR using MQSM. In this study, the use of MQSM evaluated using the densities of a given molecule computed in water and 1-octanol environments is reported as a new, completely theoretical descriptor of molecular hydrophobicity. The closeness of the linear relationships between log P values and MQSM suggests that the present approach provides a new, simple theoretical alternative to the log P parameter. It can be seen that future systematic exploration of this approach can contribute, not only to the rationalization of drug design, but to the theoretical foundation of empirical QSPR.

TABLE IV.
Comparison between Calculated and Observed Log P Values.

Molecule	log P_{obs}	log P[a]	log P[b]
Methane	1.09	1.19	1.09
Ethane	1.81	1.75	1.32
Propane	2.36	2.30	1.78
Butane	2.89	2.85	2.24
Pentane	3.39	3.40	2.70
Hexane	3.90	3.95	3.17
Methylamine	−0.57	−0.60	−0.50
Ethylamine	−0.13	−0.07	−0.02
Propylamine	0.48	0.45	0.44
Butylamine	0.97	0.97	0.90
Methanol	−0.77	−0.85	−0.45
Ethanol	−0.31	−0.26	0.03
Propanol	0.25	0.33	0.50
Butanol	0.88	0.90	0.96
Pentanol	1.56	1.47	1.42
Hexanol	2.03	2.05	1.88
2-Propanol	0.05	0.06	−0.05
2-Butanol	0.61	0.62	0.41
2-Pentanol	1.19	1.18	0.87
3-Pentanol	1.21	1.17	0.87
2-Hexanol	1.76	1.72	1.34
3-Hexanol	1.65	1.72	1.34

[a] Values calculated from monofunctional equations of Table II.
[b] Values calculated from the atomic contributions given by Ghose and Crippen.[7]

TABLE V.
Comparison between Calculated and Observed Log P Values.

Molecule	log P_{obs}	log P[a]	log P[b]
Propane	2.36	2.30	2.20
Pentane	3.39	3.40	3.23
Methylamine	−0.57	−0.60	−0.77
Butylamine	0.97	0.97	0.76
Methanol	−0.77	−0.85	−0.66
Ethanol	−0.31	−0.26	−0.20
Propanol	0.25	0.33	0.31
Butanol	0.88	0.90	0.82
Pentanol	1.56	1.47	1.34
Hexanol	2.03	2.05	1.85
Acetone	−0.24	−0.23	−0.20
2-Butanone	0.29	0.33	0.32
2-Hexanone	1.38	1.45	1.35
Acetic acid	−0.17	−0.23	−0.30
Methylacetate	0.18	0.24	0.29
Ethylacetate	0.73	0.75	0.74
Propionic acid	0.33	0.34	0.20
Butyric acid	0.79	0.85	0.71
Hexanoic acid	1.92	1.89	1.74
Butyramide	−0.21	−0.18	−0.40

[a] Values calculated from monofunctional equations of Table II.
[b] Values taken from Klopman et al.[6] using the charge density method.

Acknowledgments

This work has been performed during the stay of one of us (R. P.) at the University of Girona. Thanks are due to CESCA and CEPBA for the generous amounts of computation time. The authors thank the referees for their constructive criticism, which has improved many aspects of this work.

References

1. T. Fujita, J. Iwasa, and C. Hansch, *J. Am. Chem. Soc.*, **86**, 5175 (1964).
2. R. F. Rekker, *The Hydrophobic Fragmental Constants. Its Derivation and Applications. A Means of Characterizing Membrane Systems*, Elsevier, New York, 1977.
3. R. F. Rekker and H. M. Kort, *Eur. J. Med. Chem.*, **14**, 479 (1979).
4. C. Hansch and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley, New York, 1979.
5. G. Klopman and L. Iroff, *J. Comput. Chem.*, **2**, 157 (1981).
6. G. Klopman, K. Namboodiri and M. Schochet, *J. Comput. Chem.*, **1**, 28 (1985).
7. A. K. Ghose and G. M. Crippen, *J. Comput. Chem.*, **7**, 565 (1986).
8. R. Carbó, L. Leyda, and M. Arnau, *Int. J. Quantum Chem.*, **17**, 1185 (1980).
9. E. Besalú, R. Carbó, J. Mestres, and M. Solà, *Topics Curr. Chem.*, **173**, 31 (1995).
10. R. Carbó, E. Besalú, B. Calabuig, and L. Vera, *Adv. Quant. Chem.*, **25**, 253 (1994).
11. R. Carbó, Ed., *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, Kluwer, Amsterdam, 1995.
12. R. Carbó-Dorca and P. G. Mezey, Eds., *Advances in Molecular Similarity, Vol. 1*, JAI Press, Greenwich, CT, 1996.
13. J. Cioslowski and E. D. Fleischmann, *J. Am. Chem. Soc.*, **113**, 64 (1991).
14. C. Burt, W. G. Richards, and P. Huxley, *J. Comput. Chem.*, **10**, 1139 (1990).
15. P. G. Mezey, *Topics Curr. Chem.*, **173**, 63 (1995).
16. N. L. Allan and D. L. Cooper, *Topics Curr. Chem.*, **173**, 85 (1995).
17. R. Ponec, *Topics Curr. Chem.*, **174**, 1 (1995).
18. R. Carbó, E. Besalú, L. Amat, and X. Fradera, *J. Math. Chem.*, **18**, 237 (1995).
19. X. Fradera, L. Amat, E. Besalú, and R. Carbó-Dorca, *Quant. Struct.-Act. Relat.*, **16**, 25 (1997).
20. M. Lobato, L. Amat, E. Besalú, and R. Carbó-Dorca, *Quant. Struct.-Act. Relat.*, **16**, 465 (1997).
21. P. Constans, L. Amat, and R. Carbó-Dorca, *J. Comput. Chem.*, **18**, 826 (1997).
22. P. Constans and R. Carbó, *J. Chem. Inf. Comput. Sci.*, **35**, 1046 (1995).
23. L. Amat and R. Carbó-Dorca, *J. Comput. Chem.*, **18**, 2023 (1997).
24. R. Carbó-Dorca, *J. Math. Chem.*, **22**, 143 (1997).
25. ASA coefficients and exponents can be seen and downloaded from the WWW site: http://iqc.udg.es/cat/similarity/ASA/funcset.html
26. See for example: V. R. Saunders, In *Computational Techniques in Quantum Chemistry and Molecular Physics*, Reidel, Dordrecht, 1975, p. 347.
27. M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople, *Gaussian-94* (Revision E.2), Gaussian, Inc., Pittsburgh , PA, 1995.
28. S. Miertus, E. Scrocco, and J. Tomasi, *Chem. Phys.*, **55**, 117 (1981).
29. S. Miertus and J. Tomasi, *Chem. Phys.*, **65**, 239 (1982).
30. *Handbook of Chemisty and Physics, 63rd Ed.*, CRC Press, Boca Raton, FL, 1982.
31. J. Mestres, M. Solà, R. Carbó, F. J. Luque, and M. Orozco, *J. Phys. Chem.*, **100**, 606 (1996).
32. C. Hansch, D. Kim, A. J. Leo, E. Novellino, C. Silipo, and A. Vittoria, *CRC Crit. Rev. Toxicol.*, **19**, 185 (1989).
33. C. Hansch, A. Leo, and D. Hoekman, *Exploring QSAR, Hydrophobic, Electronic, and Steric Constants*, ACS, Washington, DC, 1995.
34. C. Hansch and A. Leo, *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*; ACS, Washington, DC, 1995.